JOURNAL OF LEARNING ANALYTICS
SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH
Volume 6(3), 25–30. http://dx.doi.org/10.18608/jla.2019.63.5

Invited Dialogue: "What's the Problem with Learning Analytics?" (Selwyn, 2019)

Ethical Challenges for Learning Analytics

Rebecca Ferguson[1]

Corresponding author [1] Email: rebecca.ferguson@open.ac.uk Address: IET, Jennie Lee Building, The Open University, Milton Keynes, MK5 8BD, UK

1. Introduction

In 2016 the Journal of Learning Analytics included a special section on Ethics and Privacy in Learning Analytics. This built on the extensive work of an ongoing series of international workshops on ethics and privacy in learning analytics (EP4LA). The associated guest editorial (Ferguson, Hoel, Scheffel, & Drachsler, 2016) reviewed the eight papers in the section and identified a series of learning analytics challenges with ethical dimensions. These challenges could be clustered under six headings: duty to act, informed consent, safeguarding, equality and justice, data ownership and protection, and privacy and integrity of self.

Those sets of challenges were identified within the learning analytics community. Selwyn's paper "What's the problem with learning analytics?" (2019) offers an external perspective. It provides an opportunity to revisit these sets of challenges, providing more detailed accounts of the areas that they cover, and identifying ways in which they can be revised in order to guide research and practice.

2. Challenges

Selwyn raises many concerns about learning analytics in his paper (2019). The following sections consider the ways in which these issues relate to the six broad areas of ethical challenge identified in the past (Ferguson, Hoel, et al., 2016).

2.1. Duty to Act

A primary responsibility of educators and educational institutions is helping their students to learn successfully. It is therefore important to make use of any available resources that will help to achieve this goal. Slade and Prinsloo (2015) suggest institutions may have "a moral responsibility for employing information which aims to provide more effective and relevant support for all students." As the purpose of learning analytics is "understanding and optimising learning and the environments in which it occurs" (Long & Siemens, 2011, p. 34), institutions have an obligation to explore the possibility that they can support their students by making use of the data generated as a by-product of learning and teaching activities.

This duty to act is not addressed by Selwyn. However, some of the issues that he raises challenge the idea that learning analytics help institutions to improve student support. He observes that many elements of education cannot be fully captured and expressed through data processing. That processing may also introduce artificial boundaries, for example by separating teaching processes from learning processes. If these constraints of data collection and processing are not recognized and accounted for, the recommendations of learning analytics will be unreliable because they are based on an impoverished or distorted view of education. Because of this, Echeverria and her colleagues have identified that "a frontline challenge is the enrichment of quantitative data streams with the qualitative insights needed to make sense of them" (Echeverria, Martinez-Maldonado, & Buckingham Shum, 2019, p. 1).

Another challenge to the duty to act relates to a concern that learning analytics are not being developed, or used, to benefit learners or teachers. Instead, they are being used to benefit institutions. This problem is associated with a blurring of the boundaries between "learning analytics — which benefit learners and faculty and are focused at the level of courses and department — and academic analytics — which benefit funders, administrators and marketing at institutional level; funders and administrators at regional level; and governments and education authorities at (inter)national level" (Ferguson, 2012). There is a noticeable gap here between the work being done by the learning analytics research community and the wide variety of analytic tools presented at educational shows for the management of staff, learners, and records. One promising route forward, which ensures that learning analytics are firmly connected to student success, is the development of student-led analytics (Evrard & Teplovs, 2017; Knox, 2017).

The "duty to act," as identified in 2016, was a heading used to bring together six challenges with ethical dimensions: use of data to benefit learners, provision of accurate and timely data, valid and reliable results of analysis, opportunities to correct data and analysis, data and results that are comprehensible to end users, and that are presented in ways that support learning.

Selwyn's paper introduces two more elements: the need to ensure that analytics take into account all that we know about teaching and learning, and the need to ensure that learning analytics are used to optimize learning.

2.2. Informed Consent

In order to optimize learning, learning analytics require the collection and analysis of data relating to individuals and groups of learners. This is an area where ethical standards and expectations are shifting relatively quickly. The notion of informed consent originated in the medical sciences, as doctors became aware that experiments could leave them open to charges of assault (Murray, 1990). In the second half of the twentieth century, the practice was taken up within the social sciences, prompted by the principle of doing no harm and the development of the 1947 Nuremberg Code. The practice was not automatically extended to data mining and analytics, in part because the potential for harm was not foreseen. More recently, it has become clear that misuse and misinterpretation of personal data pose a significant risk. As a result, requirements for informed consent are being negotiated at many levels and the area is increasingly regulated at a national or international level (for example by the General Data Protection Regulation — GDPR — implemented in Europe in 2018).

In 2016, the challenge associated with informed consent was simply that it was necessary but not easy to obtain responsibly (Ferguson, Hoel, et al., 2016). Selwyn raises a series of other pertinent issues. He notes that teaching and learning are processes that involve knowing why, as well as what. Simply being presented with the outcomes of a set of algorithms does not address this requirement, it does not strengthen learners' ability to make decisions for themselves, and it can be a disempowering experience, discouraging both learners and educators from exercising their own judgment. Analytic findings may be perceived as direct measures, rather than the proxies or indicators they actually are. In addition, machine-learning techniques make it increasingly difficult to look inside the "black box" in order to understand the rationale behind the results generated.

Some work is underway to address these challenges. For example, García and her colleagues introduced the possibility that learning analytics could provide a way of "peeking into the black box," supporting self-reflection, awareness, and decision making (García et al., 2012). More broadly, scholars are exploring the idea of "algorithmic accountability" — what this involves and how it can be implemented (Diakopoulos, 2014; Knight, Buckingham Shum, Ryan, Sándor, & Wang, 2018). The need to develop the data literacy of learners, educators, and the population at large is becoming clearer (Alhadad, 2018), with some courses in the subject now running or under development (Ferguson, Brasher, et al., 2016).

More broadly, Selwyn's critiques highlight the need for learning analytics researchers and developers to have a well-developed understanding of the processes of teaching and learning, and the wider need for increased data literacy. Students need the tools that will help them to understand what they are consenting to, and researchers/developers need to have a clear understanding of the pedagogic value that they are offering.

2.3. Safeguarding

This understanding of the value of the analytics on offer is one aspect of safeguarding. More broadly, it is a well-established ethical principle that we should safeguard those in our care and that teachers, particularly those responsible for children and young teenagers, act *in loco parentis*, in place of a parent. In the context of learning analytics, this is associated with requirements to safeguard individuals' interests and rights, to provide additional safeguards for vulnerable individuals, and to publicize mechanisms for complaint and correction of errors (Ferguson, Hoel, et al., 2016).

Selwyn points to the possibility that learning analytics will promote undesirable behaviours that work against the best interests of learners and educators. He argues that data and analytics can never fully capture the social complexity of classrooms and the complicated lives that students lead. Data may be inaccurate, incomplete, poorly chosen, or poor indicators and cannot adequately model educational processes and practices. It is possible that teachers using analytics will no longer have the best interests of their students at heart; instead, they will outsource work to algorithms that are incapable of fully understanding the learning process.

This is, of course, just one aspect of a much wider problem that extends across the field of education. Learning is rarely a visible process, and so it is necessary to work with proxies, such as assessment and exam marks, when deciding which approaches work best. Even the most talented teacher cannot fully grasp the social complexity of a single classroom, a problem exacerbated when working with multiple groups of learners.

The rise of learning analytics has not created these problems, but it has focused attention upon them. One option is to use analytics to supplement, rather than replace, human understanding, exploring how technology can augment human intelligence (Pardo, Jovanovic, Dawson, Gašević, & Mirriahi, 2019). An alternative approach is to use analytics to investigate why a learning event or program did or did not succeed, rather than to focus on the success of individual students (Liu, Rogers, & Pardo, 2015). In both cases, analytics are used to supplement, rather than replace, the capabilities of educators.

Another concern raised by Selwyn that relates to safeguarding issues is that both educators and students may use analytics to game the educational system. Educators will "teach to the algorithm" just as they teach to the test. Students will work in ways that are judged "good" by analytics indicators rather than in ways that actually help them to learn. Practice will be constrained to conform with what can be measured and analyzed. Again, the shift towards learning analytics highlights, rather than creates, the long-standing problem that proxies (exams and grades) are often considered more important than the learning they may represent. Work on reducing gaming in education predates learning analytics (for example, Arroyo et al., 2007), and is now expanding in the field to take into account aspects of equality (Albuquerque, Bittencourt, Coelho, & Silva, 2017). More broadly, learning analytics provides a growing set of tools that can inform educators about ways in which specific students are struggling, providing an insight into interactions and dialogue that was not possible in the past, and supporting an understanding of why some individuals are choosing to game the system.

2.4. Equality and Justice

Like safeguarding, equality and justice are well-recognized ethical goals, but the challenges associated with them in a learning analytics context have not yet been fully explored. In 2016, the need to share insights and findings across digital divides and a requirement to comply with the law (Ferguson, Hoel, et al., 2016) were identified, but no other challenges were noted.

Selwyn raises a series of issues related to this area: the propensity of learning analytics to entrench and deepen the status quo, to disempower and disenfranchise vulnerable groups, and to (dis)advantage some groups more than others. These are serious issues, and they relate to wider issues that are emerging globally in relation to the "datafication" of society in general and education more specifically (Williamson, 2018). However, it is worth remembering that this process works two ways. Data can be used to entrench the status quo; data can also be used to expose inequalities. Richardson's extensive body of work in this area shows, for example, that "ethnic minority students in the UK are being awarded poorer degrees for reasons that have nothing to do with their academic ability" (Richardson, 2015, p. 282). As well, UK students with unseen disabilities (such as diabetes, epilepsy, and asthma) do not do as well as those with no reported disability (Richardson, 2009). Identifying inequalities is the first step towards improving teaching and institutional support for these students.

Another issue here is self-regulated learning. Selwyn notes, "the idea of the self-responsibilized, self-determining learner advantages those individuals who are able to act in agentic, self-motivated, empowered ways" (2019). It certainly does and has always done so. A strength of learning analytics is that it offers many tools that can help learners to develop skills in this area, as the papers in the 2015 special section of this journal on self-regulated learning and learning analytics indicate (Roll & Winne, 2015). In addition, by raising awareness of the importance of self-regulation for learners, learning analytics can contribute to the development of educational environments where learners can act in self-motivated and empowered ways.

2.5. Data Ownership and Protection

When the integration of learning analytics into an educational environment is considered, issues relating to data and data protection are often the ethical challenges considered first. These challenges relate to transparency of collection and use, integration of different datasets, care and management of data, security, access, and ownership.

Selwyn approaches this set of issues from a different perspective, and focuses on data as an economic resource, the product of labour that generates value for those able to claim ownership. He points to the emergence of a global data economy that extracts economic value from the collection and processing of data. These data are passed around, recombined, and (re)processed by bots and algorithms. As a result, their origins and limitations become increasingly obscure. Nevertheless, these data are valuable resources generated by learners and educators.

Learning analytics can be viewed as benign use of a by-product that is originally of no value to its producer. Selwyn presents an alternative perspective, data harvesting as an exploitative process: a form of coerced labour.

The status of data, their ownership and value, is determined in a context that extends far beyond learning analytics, in a world where data breaches and data misuse frequently hit the headlines. However, a strong case can be made that learning analytics takes the data produced by students and transforms that data into tools that benefit those same students. Both institution and student benefit from the transaction. As Slade and her colleagues found in a survey of students, "74% of respondents indicated that they are comfortable with collection of personal data in exchange for more effective, personalized support and services" (Slade, Prinsloo, & Khalil, 2019, p. 242). The challenge is to make sure that this deal is upheld, the promised support and services do materialize, and data are not simply treated as a commodity for the highest bidder.

2.6. Privacy and Integrity of Self

The appropriate use of data is an aspect of a final set of challenges relating to privacy and integrity of the self — control over how we are seen and the ability to maintain a separation between our private selves and our public personas. Privacy can be understood as "a freedom from unauthorized intrusion: the ability of an individual or a group to seclude themselves or the information about them, and thus to express themselves selectively" (Ferguson, Hoel, et al., 2016, p. 11).

ISSN 1929-7750 (online).

Selwyn's concerns here relate to the possibility that individuals will be defined by analytics rather than viewed on their own terms. Professional judgment will be undermined, students and teachers will lose control of their work, and both staff and students will constantly be monitored. To some extent, this future is already with us, with countries shifting their educational systems based on flawed interpretations of data; schools and teachers feeling constrained to achieve unrealistic targets; and students driven to suicide because of misreported exam results (Saltelli, 2017; Abrams, 2017; Wallen, 2019).

One response is that this dystopian vision relates to the misuse of academic analytics, rather than learning analytics. It applies to league tables, PISA scores, and management systems — tools that use data for administration and governance in an educational context, rather than as a direct support for learning and teaching. However, this response sidesteps the issue by drawing a boundary between the two sets of analytic. They are designed with different uses in mind, but the context remains the same, the possibilities for misuse are very similar, and most people would not be able to distinguish between academic and learning analytics.

This is the set of challenges that the field of learning analytics has done least to address. We may note in keynotes and workshops that we are faced with a fast-growing problem, but it is not a problem that we have yet begun to tackle, perhaps because it relates to deeply entrenched uses of data and analysis; perhaps because major cultural shifts are required. There is work to be done here to establish the dimensions and scale of the problem, establish the extent to which learning analytics is implicated, and develop a plan of action.

3. Discussion

In his paper, Selwyn identifies more than 30 concerns that emerge as learning analytics are implemented. All of these can be related to the six broad groups of ethical challenges already identified by the field: duty to act, informed consent, safeguarding, equality and justice, and privacy and integrity of self. This suggests that researchers and practitioners have worked together successfully in order to identify these challenges. In some cases, learning analytics can provide a partial solution to a longstanding problem; in other cases, work has already begun on these challenges.

Although the broad areas of challenge have been identified, the perspectives put forward by Selwyn expand and inform understanding of these issues. This reflects the value of seeking expert opinion from outside the field. It also indicates the value of revisiting these issues, as learning analytics evolve, and the wider context of education and data shifts.

This new consideration of the ethical challenges faced by learning analytics means that the initial versions can be revised and expanded to reflect a broader range of issues, and to indicate more clearly what needs to be done to address them.

Challenge one: Use data and analytics whenever they can contribute to learner success, ensuring that the analytics take into account all that is known about learning and teaching

Challenge two: Equip learners and educators with data literacy skills, so they are sufficiently informed to give or withhold consent to the use of data and analytics

Challenge three: Take a proactive approach to safeguarding in an increasingly data-driven society, identifying potential risks, and taking action to limit them.

Challenge four: Work towards increased equality and justice, expanding awareness of ways in which analytics have the potential to increase or decrease these.

Challenge five: Increase understanding of the value, ownership, and control of data.

Challenge six: Increase the agency of learners and educators in relation to the use and understanding of educational data.

The learning analytics community is already addressing each of these challenges, to some extent. Selwyn's paper makes it clear that there is more to be done. There is room here for work on an individual level and at a small scale. There is also a need for more substantial projects. In particular, the need for widespread data literacy in a world that is saturated with data is becoming particularly acute. This is not the sole responsibility of our community, but we are well placed to take a lead in this area.

The need to increase data literacy points to the need for learning analytics as a field to pay more attention to context. Building a learning analytics tool is just one aspect of implementation. Work is also required to ensure that the tool can be used effectively by learners and teachers, that it helps learners to learn more successfully, and that the six ethical challenges of learning analytics are considered and addressed.

4. Conclusion

This paper is one of several responses to Neil Selwyn's paper in this issue: "What's the problem with learning analytics?" It considers the problems raised by Selwyn from an ethical perspective and relates these to a long-standing strand of work in the field that deals with the ethics of learning analytics. As a result, the paper identifies and formulates six ethical challenges for learning analytics. The intention, as with the related papers in this issue, is to provoke discussion, reflection, and action within the learning analytics community.

Declaration of Conflicting Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author declared no financial support for the research, authorship, and/or publication of this article.

References

Abrams, F. (2017). Sats tests: Why cheating and cramming mean the numbers don't add up. *The Guardian*, 6 June 2017. https://www.theguardian.com/education/2017/jun/06/sats-tests-cheating-cramming-primary-school-exam

Albuquerque, J., Bittencourt, I. I., Coelho, J. A. P. M., & Silva, A. P. (2017). Does gender stereotype threat in gamified educational environments cause anxiety? An experimental study. *Computers & Education, 115*, 161–170. https://dx.doi.org/10.1016/j.compedu.2017.08.005

Alhadad, S. S. J. (2018). Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *Journal of Learning Analytics, 5*(2), 60–85. http://dx.doi.org/10.18608/jla.2018.52.5

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S. & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. *Artificial Intelligence in Education, 158*, 195–202. https://scholarworks.umass.edu/cs_faculty_pubs/326

Diakopoulos, N. (2014). Algorithmic accountability reporting: On the investigation of black boxes. Tow Center for Digital Journalism, Columbia Journalism School. https://dx.doi.org/10.7916/D8ZK5TW2

Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019). Towards collaboration translucence: Giving meaning to multimodal group data. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 4–9 May 2019, Glasgow, Scotland, UK, Paper No. 39. New York: ACM. http://dx.doi.org/10.1145/3290605.3300269

Evrard, A. E., & Teplovs, C. (2017). Community building around a shared history: Rebooting academic reporting tools at the University of Michigan. Paper presented at the 7th International Conference on Learning Analytics and Knowledge (LAK '17), 13–17 March 2017, Vancouver, BC, Canada (practitioner track).

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. https://dx.doi.org/10.1504/IJTEL.2012.051816

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T. & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. Joint Research Centre Science for Policy Report, EUR 28294 https://dx.doi.org/10.2791/955210

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Ethics and privacy in learning analytics. *Journal of Learning Analytics, 3*(1), 5–15.

García, R. C., Pardo, A., Delgado Kloos, C., Niemann, K., Scheffel, M., & Wolpers, M. (2012). Peeking into the black box: Visualising learning activities. *International Journal of Technology Enhanced Learning, 4*(1/2), 99–120. http://dx.doi.org/10.1504/IJTEL.2012.048313

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education, 28*(1), 1–28. doi: https://dx.doi.org/10.1007/s40593-016-0121-0

Knox, J. (2017). Playing with student data: The Learning Analytics Report Card (LARC). Paper presented at the 7th International Conference on Learning Analytics and Knowledge (LAK '17), 13–17 March 2017, Vancouver, BC, Canada (practitioner track).

Liu, D. Y.-T., Rogers, T., & Pardo, A. (2015). Learning analytics: Are we at risk of missing the point? In T. Reiners, B. R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), *Globally Connected, Digitally Enabled: Proceedings of the 32nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2015),* 29 November–2 December 2015, Perth, Western Australia (pp. 684–687). Australasian Society for Computers in Learning in Tertiary Education.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.

Murray, P. M. (1990). The history of informed consent. *The Iowa Orthopaedic Journal, 10*, 104–109.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128–138. doi: https://dx.doi.org/10.1111/bjet.12592

Richardson, J. T. E. (2009). The academic attainment of students with disabilities in UK higher education. *Studies in Higher Education, 34*(2), 123–137. https://dx.doi.org/10.1080/03075070802596996

Richardson, J. T. E. (2015). The under-attainment of ethnic minority students in UK higher education: What we know and what we don't know. *Journal of Further and Higher Education, 39*(2), 278–291. doi: https://dx.doi.org/10.1080/0309877X.2013.858680

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self–regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7–12. https://dx.doi.org/10.18608/jla.2015.21.2

Saltelli, A. (2017). International PISA tests show how evidence-based policy can go wrong. *The Conversation*, 12 June 2017. https://theconversation.com/international-pisa-tests-show-how-evidence-based-policy-can-go-wrong-77847

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics, 6*(3), 11–19. http://dx.doi.org/10.18608/jla.2019.63.3

Slade, S., & Prinsloo, P. (2015). Student perspectives on the use of their data: Between intrusion, surveillance and care. *European Journal of Open, Distance and E-learning, 18*(1). https://www.eurodl.org/?p=special&sp=articles&inum=7&article=679

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (LAK '19), 4–8 March 2019, Tempe, Arizona, USA (pp. 235–244). New York: ACM. http://dx.doi.org/10.1145/3303772.3303796

Wallen, J. (2019). Over twenty Indian students commit suicide after inaccurate university admission results. *The Telegraph,* April 30, 2019. https://www.telegraph.co.uk/news/2019/04/30/twenty-indian-students-commit-suicide-inaccurate-university/

Williamson, B. (2018, March 17). 10 definitions of datafication (in education). https://codeactsineducation.wordpress.com/2018/03/17/10-definitions-datafication/